



Comparison of modeling techniques for milk-production forecasting

M. D. Murphy,*† M. J. O'Mahony,† L. Shalloo,* P. French,* and J. Upton*¹

*Animal and Grassland Research Innovation Centre, Teagasc Moorepark, Co. Cork, Ireland

†Department of Mechanical Engineering, Cork Institute of Technology, Co. Cork, Ireland

ABSTRACT

The objective of this study was to assess the suitability of 3 different modeling techniques for the prediction of total daily herd milk yield from a herd of 140 lactating pasture-based dairy cows over varying forecast horizons. A nonlinear auto-regressive model with exogenous input, a static artificial neural network, and a multiple linear regression model were developed using 3 yr of historical milk-production data. The models predicted the total daily herd milk yield over a full season using a 305-d forecast horizon and 50-, 30-, and 10-d moving piecewise horizons to test the accuracy of the models over long- and short-term periods. All 3 models predicted the daily production levels for a full lactation of 305 d with a percentage root mean square error (RMSE) of $\leq 12.03\%$. However, the nonlinear auto-regressive model with exogenous input was capable of increasing its prediction accuracy as the horizon was shortened from 305 to 50, 30, and 10 d [RMSE (%) = 8.59, 8.1, 6.77, 5.84], whereas the static artificial neural network [RMSE (%) = 12.03, 12.15, 11.74, 10.7] and the multiple linear regression model [RMSE (%) = 10.62, 10.68, 10.62, 10.54] were not able to reduce their forecast error over the same horizons to the same extent. For this particular application the nonlinear auto-regressive model with exogenous input can be presented as a more accurate alternative to conventional regression modeling techniques, especially for short-term milk-yield predictions.

Key words: dairy production, milk-production forecasting, modeling

INTRODUCTION

Milk production from pasture-based dairy cows is susceptible to variation due to seasonality of pasture production (Adediran et al., 2012), grazing conditions (Baudracco et al., 2012), disease (Collard et al., 2000), nutritional interventions (Kolver and Muller, 1998), and other disturbances (Olori et al., 1999; Tekerli et

al., 2000). The ability to forecast herd milk yield days, weeks, and months in advance provides benefits for management at processor and farm level as total daily milk production strongly influences energy consumption, plant utilization, and farm income. The usefulness of a milk-yield prediction system depends upon how accurately it can predict daily milking patterns and its ability to adjust to factors affecting supply. Milk yield prediction models have proven useful for genetic analysis (Ptak and Schaeffer, 1993) and for bio-economic modeling (Shalloo et al., 2004).

Studies have been undertaken by Wood (1967), Ali and Schaeffer (1987), Wilmink (1987), and Guo (1995), who all developed algebraic equations for the purpose of fitting a lactation curve to empirical data. Jones (1997) stressed the need for increased flexibility and adaptation among curve-fitting techniques and introduced an empirical Bayes method for fitting Wood's lactation curve (incomplete gamma function; Wood, 1967). Macciotta et al. (2002) and Vasconcelos et al. (2004) employed auto-regressive models to predict individual lactations using limited numbers of test days throughout the lactation cycle. Other attempts to forecast milk yields have involved large regression models such as artificial neural networks (**ANN**) and multiple linear regression (**MLR**) models (Lacroix et al., 1995; Salehi et al., 1998; Sharma et al., 2006; Sharma et al., 2007). These models proved to be very successful; however, they require large amounts of detailed information for each specific cow. The ANN model developed by Sharma et al. (2007) requires 12 individual traits of each cow (genetic group, season of birth, period of birth, birth weight, age at maturity, weight at maturity, season of calving, period of calving, age at calving, weight at calving, peak yield, and days to attain peak yield); likewise the model tested by Lacroix et al. (1995) required 16 network inputs including information such as logarithm of somatic cell count, energy fed on test day, protein fed on test day, DM fed on test day, and so on. Brun-Lafleur et al. (2010) modeled variation in milk yield with respect to energy and protein supply, but acquiring even this information for an entire pasture-based herd is not practical. A balance is required between the availability of detailed

Received September 2, 2013.

Accepted February 11, 2014.

¹Corresponding author: john.upton@teagasc.ie

information at farm level and the prediction accuracy of the milk-supply model. The high levels of detailed data required to construct these milk-yield predictors inhibit their practical implementation on commercial dairy farms.

The aim of this study was to assess the suitability of a static neural network (**SANN**), a MLR model, and a nonlinear auto regressive model with exogenous input (**NARX**) for the prediction of total daily herd milk yield (**DHMY**) over varying forecast horizons. The most successful model was selected according to its abilities to generate the most accurate forecast using very limited training data in low volumes over a long- (305 d), medium- (30 to 50 d), and short-term (10 d) horizon.

MATERIALS AND METHODS

Data Collection

Data were collected from a research farm in the south of Ireland for a period of 4 yr (2006–2010). Daily herd milk yield (liters) and number of cows milked (**NCM**) on that corresponding herd DIM was collected because this is the most accessible data for commercial farmers to obtain. Milk yields were recorded from a conventional herringbone swingover milking parlor using ICAR (International Committee for Animal Recording)-approved milk meters. The model was set at herd level and evaluated by comparing daily milk yields across a herd of 140 pasture-based Holstein-Friesian (**HF**) cattle (North American HF and New Zealand HF genetic strains). The milking season of 2010 was selected as the target prediction horizon and the previous 3 yr of data were used to train the model.

Model Inputs

In previous studies certain variables were found to have an influence on milk production: season of calving (Wood, 1967), climatic conditions (Smith, 1968), number of DIM (Grzesiak et al., 2006), and stocking rate (McCarthy et al., 2011). In this study the farm grazing area remained static, whereas the number of cows grazing varied throughout the year. Similarly, the season of calving (spring) was kept constant in the herd over several years. Hence the total herd milk production behaves in a cyclical pattern (assuming no catastrophic external factors). This pattern is influenced by the herd size at any one time, the DIM of the herd, and other factors such as atmospheric conditions (ambient temperature, irradiance, and precipitation). This information is readily available on commercial dairy farms; hence, DIM and NCM were selected as model inputs.

The localized prediction of atmospheric conditions was deemed outside the scope of this study.

The models were trained with basic information (DIM and NCM) and used to predict DHMY over specified time horizons. The total DHMY can be viewed as a time series that is being primarily driven by a handful of factors. The number of cows coming in and out of lactation can be factored in by recording the NCM on each milking day. The DIM is factored in by chronologically applying a day number relative to the beginning of lactation for the entire herd. All 3 model predictions were trialed over several prediction horizons: 305, 50, 30, and 10 d. For the horizons less than 305 d, the models repeatedly projected over the specific horizon in a moving piecewise manner until the end of the series. After every horizon step the previous DHMY data were added to the models training set before the next prediction, updating the model state. (All 3 models were developed using the software package MATLAB R2012a; Mathworks, Natick, MA.) The statistics toolbox was used to create the MLR model, and the neural networks toolbox was used to create the neural network models. For detailed information regarding the data processing, structure, and training of these models, please refer to Demuth et al. (2010).

Neural Networks

An ANN is a mathematical model whose operating principle is based on biological neural networks (Haykin, 1999). The ANN architecture comprises a series of interconnected layered neurons through which inputs are processed. These inputs values are multiplied by the synaptic weights, which represent the strength of the neural connections. Figure 1 shows a typical feedforward ANN structure containing an input, hidden, and output layer. This configuration is very popular for function approximation in systems where no time-dependent relationship exists among the network inputs. Increasing the size of the hidden layer allows for more intricate function fitting of nonlinear processes; however, overfitting of training data is undesirable when good generalization abilities are needed (Demuth et al., 2010). Many methods exist for improving generalization such as data filtering, feedback elements, regularization, and network reduction. Reducing the number of neurons in the hidden layer is an effective method of improving generalization because small networks do not have the capability of overfitting the training data. The synaptic weights are configured during back propagation training (Hecht-Nielsen, 1989). Once trained, a SANN has no feedback elements and contains no delays.

Dynamic artificial neural networks are also known as recurrent neural networks because of their dynamic

Input Nodes

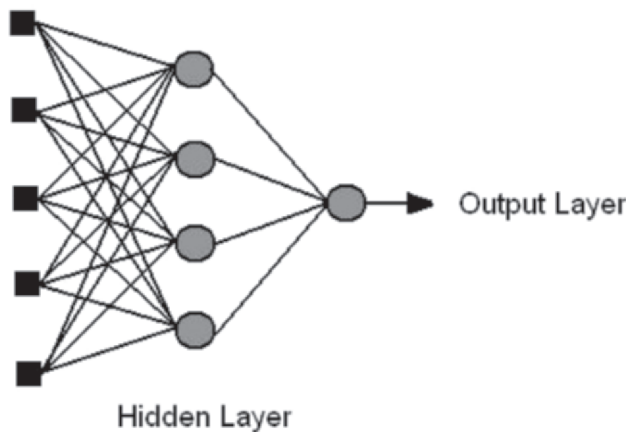


Figure 1. Neural network structure.

feedback elements, which conventional static feedforward networks do not possess. Whereas static feedforward networks generate an output based on the current model inputs, dynamic artificial neural networks produce an output based on the current inputs plus the previous inputs and outputs over a retrospective time range. The purpose of using input and output data from recent time steps is to enable short-term memory and pattern recognition within the network. Tapped delay lines (**TDL**) are placed before the input $u(t)$ and the feedback loop from the output $y(t)$ entry to the hidden layer (Figure 2). Tapped delay lines are used to store previous values of the input series $u(t)$ and target series $y(t)$ [$y(t-1)$ $y(t-2)$... $y(t-d)$], where d is the number of delays, giving the model embedded memory. It has been shown that static feedforward networks with input delays possess temporary pattern-recognition abilities for chaotic time series (Lapedes and Farber, 1987; Weigend et al., 1990).

Dynamic networks with an exogenous feedback element (from outside the network layers) and TDL are referred to as NARX models. The NARX models have proven to be a very successful modeling tool for non-linear systems and especially time series (Diaconescu, 2008). The NARX networks converge faster, are less likely to develop long-term dependencies, and typically have better generalization abilities than other networks (Tsunghan et al., 1996).

To create an ANN with function approximation or pattern-recognition abilities, the model parameters (synaptic weight and biases) must be manipulated to produce the desired network state. The actual output signal y_{pm} (for pattern p at output m) from a neuron in the output layer is driven by the signal vector \mathbf{x}_{pm} produced from the neurons in the hidden layer. The error

signal e_{pm} is the difference between the actual output y_{pm} and the desired output d_{pm} :

$$e_{pm} = d_{pm} - y_{pm}. \quad [1]$$

The error signal e_{pm} activates a training algorithm, which modifies the ANN parameters sequentially to minimize the error function derived from the error signal e_{pm} (typically the sum of the square error):

$$E_D = \sum_{p=1}^P \sum_{m=1}^M e_{p,m}^2, \quad [2]$$

where E_D is the sum of the square network errors, P is the number of network patterns and p is the pattern index, and M is the number of outputs and m is the output index. A training algorithm calculates how the network parameters need to be modified to minimize the error function; once the network parameters have been initially modified, the sequence is repeated until the cost function minimizes. In the case of multiple patterns, the cost function is calculated as the combined sum of all pattern errors. The value of E_D relative to the model parameters variation for all patterns can be visualized as an error surface. The change in E_D relative to the change in network parameters is the slope of the error surface during that learning step. The training algorithm follows in the direction of the negative gradient at each step until the lowest point on the error surface is reached.

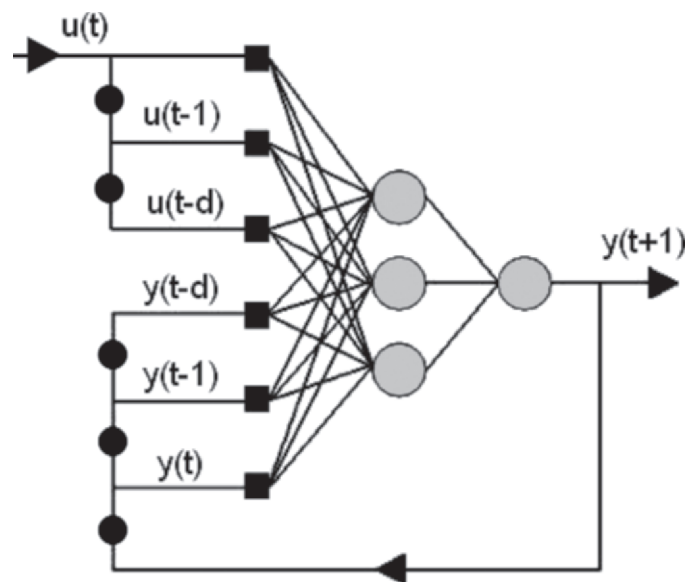


Figure 2. Network with tapped delay lines. $u(t)$ = input series; $y(t)$ = output series, where d is the number of delays.

SANN Configuration

In this study the most suitable SANN architecture was found heuristically by testing combinations of number of neurons, transfer functions, and training algorithms. The selection criterion for the best network configuration was the average root mean square error (RMSE) of the 4 prediction horizons over the test year. The optimum network consisted of 4 neurons in the hidden layer. It was observed that with the fewer number of neuron used the model was less liable to overfit. This is desirable because good generalization properties are required. A log-sigmoid and linear transfer function was used in the hidden and output layers. Training data were normalized by setting the minimum value to -1 and the maximum value to 1 ; the data in between these limits were rescaled accordingly. The model training data set (4 yr of milk production) was divided into 3 subsets (network training, validation, and testing), and the data allocation for each data set was 90, 5, and 5%, respectively. The network training set was used for gradient calculation and weight updating, the validation set was used to monitor the error during training, and the test set was used for initial testing. All 3 subsets were divided up randomly from the model training data set (Demuth et al., 2010). The Levenberg-Marquardt (Hagan and Menhaj, 1994) back-propagation algorithm was found to be the most effective training algorithm. The synaptic weight update rule for the Levenberg-Marquardt training algorithm can be written as

$$w_{k+1} = w_k - (\mathbf{J}^T \mathbf{J} + \varsigma \mathbf{I})^{-1} \mathbf{J}_k \mathbf{e}_k, \quad [3]$$

where w is the synaptic weight, k is the training iteration number, \mathbf{J} is the Jacobian matrix, \mathbf{e} is the error vector, ς is the Levenberg damping factor, \mathbf{I} is the identity matrix, and $\mathbf{J}^T \mathbf{J}$ is the approximation of the Hessian matrix. Newton's method relies on calculating the Hessian matrix to iteratively update the synaptic weights in the network. However, this involves calculating the second-order derivatives of the error function E_D . The Gauss-Newton training algorithm bypasses calculating the second-order derivatives of E_D by approximating the Hessian matrix with $\mathbf{J}^T \mathbf{J}$. However, the Gauss-Newton training algorithm may face convergence problems for complex error surfaces. In certain cases the Jacobian approximator $\mathbf{J}^T \mathbf{J}$ may not be invertible. The Levenberg-Marquardt algorithm ensures invertability of the Jacobian approximator $\mathbf{J}^T \mathbf{J}$ by introducing an identity matrix \mathbf{I} and the Levenberg damping factor ς . The size of Levenberg damping factor ς is adjusted at each iteration, which guides the training process. If the reduction in E_D between each iteration is rapid, the

value of Levenberg damping factor ς is reduced and the update rule approaches the Gauss-Newton method; if the change in the cost function is insufficient, ς is increased giving an update closer to the gradient descent direction.

Dynamic ANN Configuration

The NARX model was trained in a series parallel architecture. It has been observed that forecasting of a time series with NARX models will be enhanced by simultaneously analyzing related time series (Diaconescu, 2008). The NARX in this study was trained using DHMY as the predicted time series with DIM and NCM fed in as corresponding time series. The final NARX configuration consisted of 2 neurons in the hidden layer, 2 delays in the TDL, a tan-sigmoid and linear transfer functions in the hidden layer and output layer, respectively. The best training style was found to be Bayesian regularization. Regularization aims to keep the network weights as small as possible to improve generalization. To generalize well, a network has to be able to fit the training data while responding accurately to new data. The objective of the above training algorithms is to minimize the error function E_D . However, in the case of regularization training an extra term is added based on the square of the synaptic weights:

$$E_w = \sum_{p=1}^P \sum_{m=1}^M w_{p,m}^2. \quad [4]$$

The new training cost function F for regularization can be expressed as

$$F = \alpha E_D + \beta E_w, \quad [5]$$

where E_D is the sum of the square network errors (Equation 2), E_w is the sum of the square network weights, and α and β are objective function parameters. For the network to properly fit the data, E_D must be minimized; however, for the network to generalize well, the weights must be small. The balance between error minimization and weight reduction is controlled by the objective function parameters α and β . The relative size of the objective function parameters governs the training emphasis. If $\alpha \gg \beta$, then the training will focus on error reduction and the network will be more inclined to overfit the training data. If $\alpha \ll \beta$, then the network weights will be reduced at the cost of network fitting, thus a smoother network output is achieved for novel data (Dan Foresee and Hagan, 1997). Selection of the correct values for α and β is achieved by using Bayes'

theorem (MacKay, 1992). The regularization training in this study used the Levenberg-Marquardt update method.

The NARX network produced an output one step ahead of the input data; if another step was required, then the previous output was fed back in a closed loop to predict the next step again. This process was repeated as many times as was necessary. In short the model made a prediction that was influenced by its own previous prediction and not real data. This was not the case, however, with the SANN and MLR models because they could not store previous predictions.

MLR Model

Linear regression models characterize the association between a dependent and independent variable. The relationship between these variables can be expressed in a 2-dimensional space. However, few outputs can be accurately profiled using just one input. Most real-world systems are controlled by multiple inputs. Such systems can be expressed in the form of a MLR model:

$$Y = \varepsilon + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n, \quad [6]$$

where Y is the dependent variable, ε is a residual error, X_1, X_2, \dots, X_n are the independent variables, and $\alpha_1, \alpha_2, \dots, \alpha_n$ are the regression coefficients. In this study, DHMY was the dependent variable, with DIM and NCM as the independent variables. The relationship among DHMY, DIM, and NCM exists in a 3-dimensional mathematical space with each variable plotted on independent axes. The regression coefficients and error residual of the MLR model (Equation 8) are updated as the historical data set changes because of the shortening of the prediction horizon:

$$\text{DHMY} = \varepsilon + \alpha_1 \text{NCM} + \alpha_2 \text{DIM}. \quad [7]$$

Model Evaluation Tools

The following statistical methods have been employed in cognate studies to evaluate milk-yield prediction models (Jones, 1997; Olori et al., 1999; Grzesiak et al., 2003, 2006; Quinn et al., 2005; Sharma et al., 2006). Analyzing the model predictions using multiple techniques allows for an easier comparison of results across studies.

Coefficient of Determination. The R^2 value (Equation 3) indicates the goodness of fit between the predicted production levels and the actual values. For the i th record, ε is the residual error term, y is the predicted value, and \bar{y} is the mean of the actual production values:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\varepsilon_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad [8]$$

Variance. The variance (σ^2 ; Equation 4) is a measure of the distribution of the errors around the mean. Variance indicates spread of the prediction errors.

$$\sigma^2 = \frac{\sum_{i=1}^N (\varepsilon_i - \mu_i)^2}{N}, \quad [9]$$

where μ is the mean residual error and N is the total number of records.

RMSE. Unlike the mean absolute error (MAE), which weights the error residuals linearly, the RMSE (Equation 10) squares the residual errors before averaging, thus a quadratic weighting is applied to the error value. This makes RMSE a useful indicator of large residual outliers.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\varepsilon_i)^2}{N}}. \quad [10]$$

RESULTS

305-d Prediction Horizon

Figure 3 shows the NARX, SANN, and MLR models DHMY prediction over a full cycle (305 d). The NARX model had the highest coefficient of determination ($R^2 = 0.936$); the MLR ($R^2 = 0.917$) and SANN ($R^2 = 0.889$) models were less accurate. The RMSE of the NARX prediction (8.59%) was lower than that of the SANN (12.03%) and MLR (10.62%), indicating the presence of larger residual errors in the SANN and MLR forecasts. The difference between RMSE and MAE was due to variations in the residual error profile as accuracy of the models fluctuated over the forecast horizon. The 3 models followed a similar trajectory until the peak production period began; the SANN underestimated the maximum production level by 4.8%. The NARX and MLR underpredicted the peak milk production by 12.2 and 14.7%, respectively.

50-d Prediction Horizon

The forecast prediction errors of all 3 models over a 50-d moving piecewise horizon can be seen in Table 1 and Figure 4. Each model predicted 50 d ahead, after each prediction the actual DHMY data over the previous horizon was added to the model training data and a new 50-d prediction was generated. The maximum and

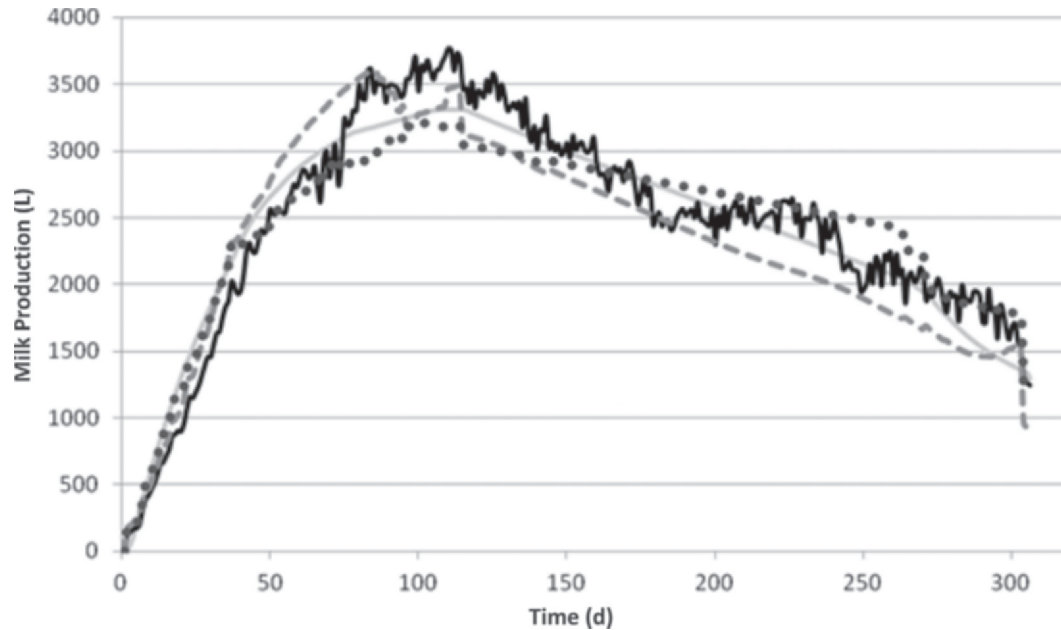


Figure 3. Forecast for 305-d milk production. Black solid line = actual milk production, gray solid line = nonlinear auto-regressive model with exogenous input forecast, dashed line = static artificial neural network forecast, dotted line = multiple linear regression forecast.

minimum RMSE values were 265.3 and 107.1 for the NARX, 389.3 and 188.6 for the SANN, and 377.2 and 194.1 for the MLR. The NARX model was the most accurate over the majority of horizons except the first (0–50 d) and the final (250–305 d) time horizon (60% of the cycle). The SANN and MLR produced similar levels of accuracy until the end of the cycle, where the MLR generated the most accurate forecast. The MLR was more accurate than the SANN in the ranges 1 to 50 d and 200 to 305 d (50% of the cycle). The smaller σ^2 values of the NARX predictions point to a more steady-state error in comparison with the other models. The NARX, SANN, and MLR underestimated the peak

milk production by 7.4, 5.9, and 14.6% of peak value, respectively.

30-d Prediction Horizon

Figure 5 shows predictions from the 3 models over a 30-d piecewise horizon. Over the 30-d piecewise intervals the NARX changed its trajectory, resulting in smaller MAE, RMSE, and σ^2 values than the other models over the cycle except ranges 0 to 30 d, 210 to 240 d, and >270 d (70% of the cycle; Table 2). The MLR was more accurate than the SANN over the ranges 30 to 60 d, 120 to 240 d, and >270 d (60% of the cycle).

Table 1. The 50-d moving piecewise horizon forecast errors for 3 milk-prediction models

Range	NARX ¹			SANN ²			MLR ³		
	MAE ⁴	σ^2 ⁵	RMSE ⁶	MAE	σ^2	RMSE	MAE	σ^2	RMSE
1–50	243.8	10,941.4	265.3	188.5	12,149.7	218.4	184.2	23,509.5	212.3
50–100	211.6	12,161.2	238.6	271.6	25,884.6	315.7	269.9	36,631.0	329.8
100–150	136.7	8,150.1	163.8	285.9	22,554.3	323.0	349.6	20,515.9	377.2
150–200	89.1	3,542.5	107.1	161.6	9,475.0	188.6	170.5	11,670.5	201.3
200–250	124.7	5,964.4	146.7	365.0	18,339.1	389.3	171.3	20,552.3	222.4
250–305	189.4	13,597.7	222.4	306.1	10,259.9	322.4	153.2	14,505.5	194.1

¹NARX = nonlinear auto-regressive model with exogenous input.

²SANN = static artificial neural network model.

³MLR = multiple linear regression model.

⁴MAE = mean absolute error.

⁵ σ^2 = variance.

⁶RMSE = root mean square error.

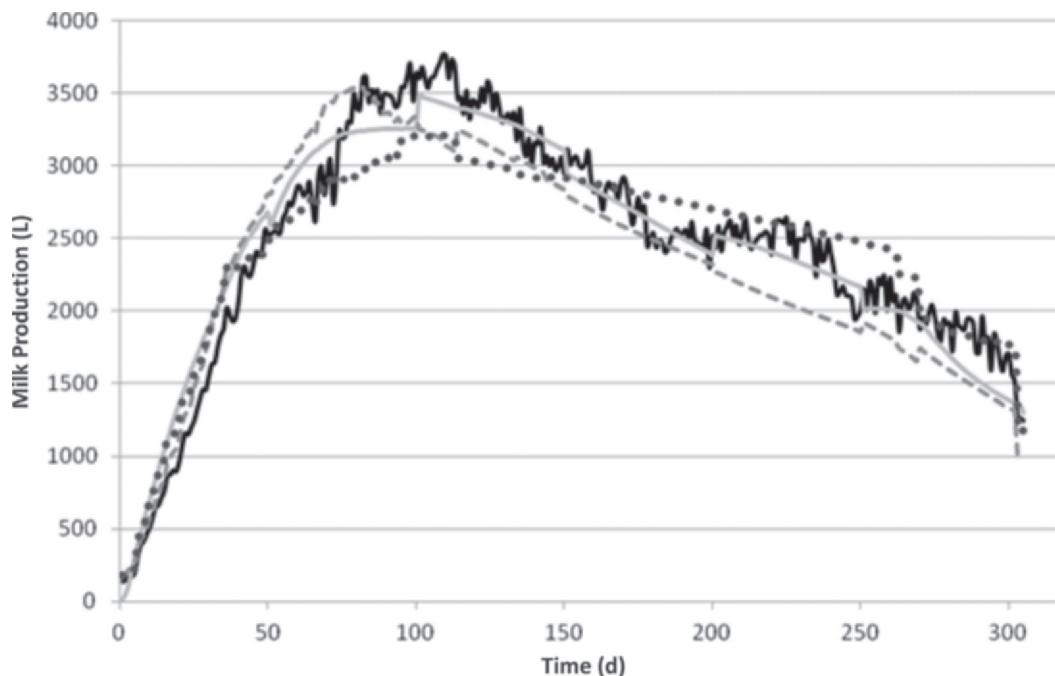


Figure 4. Milk-production forecast for a 50-d moving piecewise horizon. Black solid line = actual milk production, gray solid line = non-linear auto-regressive model with exogenous input forecast, dashed line = static artificial neural network forecast, dotted line = multiple linear regression forecast.

The maximum and minimum RMSE values were 275.2 and 77.6 for the NARX, 403.5 and 177 for the SANN, and 423.2 and 107.5 for the MLR. The NARX, SANN, and MLR peak production prediction errors were 7.5, 7, and 14.5, respectively.

10-d Prediction Horizon

Table 3 and Figure 6 show the forecast errors of the 3 models over a 10-d horizon. The NARX was the most accurate model with lower MAE, RMSE, and σ^2 values than the SANN and MLR for 83 and 73% of the forecasts, respectively. The MLR was more accurate than the SANN over 50% of the 10-d horizon forecasts. The maximum and minimum RMSE values were 213.6 and 53.9 for the NARX, 429.9 and 46.7 for the SANN, and 495.5 and 64.3 for the MLR. The NARX, SANN, and MLR peak production prediction errors were 3.9, 7.4, and 14.5% of peak value, respectively.

Model Comparisons

Table 4 shows the totalized forecast errors of the 3 models over the full 305-d cycle with 4 different piecewise horizons ranging from 305 to 10 d. The R^2 of the NARX forecast increased from 0.936 to 0.968 and the RMSE dropped from 8.59 to 5.84%, and the R^2 of the SANN forecast increased from 0.889 to 0.911 and the RMSE decreased from 12.03 to 10.7%, yet the

error of the SANN did not necessarily decrease in correspondence with a shorter prediction horizon. From the 305- to 10-d horizon, the accuracy of the MLR did not substantially increase; the R^2 and RMSE changed from 0.917 to 0.916 and 10.62% to 10.54%, respectively. The NARX achieved lower MAE, RMSE, and residual variance values for 30 of the 47 (63.8%) individual prediction ranges, the SANN was the most accurate model over 5 ranges (12.8%), and the MLR was the most accurate model over 11 ranges (23.4%). The NARX was the most accurate at predicting the entire cycle for every moving horizon; however, the MLR forecasts were more accurate for the beginning and end of the cycle. Initially, the SANN was the most successful at predicting the peak milk production with a 4.8% error in the 305-d forecast; however, the SANN peak prediction error increased to 7.4%, whereas in contrast the NARX peak prediction error reduced from 12.2 to 3.9%. The MLR peak production error varied from 14.7 to 14.5%. The NARX totalized prediction error (Table 4) dropped incrementally from 8.59 to 5.84% as the moving piecewise horizon was shortened, but the SANN and MLR error did not substantially reduce in accordance.

DISCUSSION

Model Evaluation

The NARX forecast adapted dynamically to variations in milk production as it was influenced by the

Table 2. The 30-d moving piecewise horizon forecast error

Range	NARX ¹			SANN ²			MLR ³		
	MAE ^{4,5}	σ^2 ⁶	RMSE ^{5,7}	MAE	σ^2	RMSE	MAE	σ^2	RMSE
0–30	247.0	14,689.0	275.2	149.7	17,334.2	199.4	190.7	8,916.6	212.8
30–60	62.4	2,126.2	77.6	335.8	13,267.0	355.0	138.1	12,547.9	177.8
60–90	186.0	12,394.9	216.8	221.5	29,249.5	279.8	297.8	38,576.2	356.7
90–120	172.2	13,222.3	207.1	237.9	7,187.5	252.5	415.1	6,745.9	423.2
120–150	84.1	3,489.0	102.8	304.8	8,592.9	318.6	270.7	23,278.5	310.7
150–180	103.8	6,499.2	131.5	156.4	6,862.1	177.0	114.4	7,618.3	143.9
180–210	72.1	2,466.9	87.6	204.1	15,227.5	238.5	220.1	7,622.7	236.8
210–240	129.8	6,471.6	152.7	392.2	8,991.1	403.5	99.8	5,552.8	124.5
240–270	124.1	8,220.4	153.7	192.4	11,985.8	221.3	311.5	14,558.7	334.1
>270	110.8	2,677.9	132.9	322.9	10,132.1	368.5	91.5	2,531.8	107.5

¹NARX = nonlinear auto-regressive model with exogenous input.²SANN = static artificial neural network model.³MLR = multiple linear regression model.⁴MAE = mean absolute error.⁵Error values are represented in liters.⁶ σ^2 = variance.⁷RMSE = root mean square error.

DHMY data in the previous prediction horizon. The embedded memory created by the TDL enabled the NARX model to correct its projected trajectory over the next horizon based on information from the previous horizon. The constant reduction in the NARX prediction errors was due to higher model refresh frequencies allowing the NARX model to realign its trajectory more regularly. The shorter the prediction horizon, the

fewer closed loop prediction steps the model must make, therefore limiting the effect of stochastic disturbances. The SANN and MLR models did not possess the ability to dynamically learn from their errors. The SANN prediction error did fluctuate and reduce slightly, but this was due to changes in the state of the model as it retrained before each new prediction. A very small variation existed in the MLR prediction over different

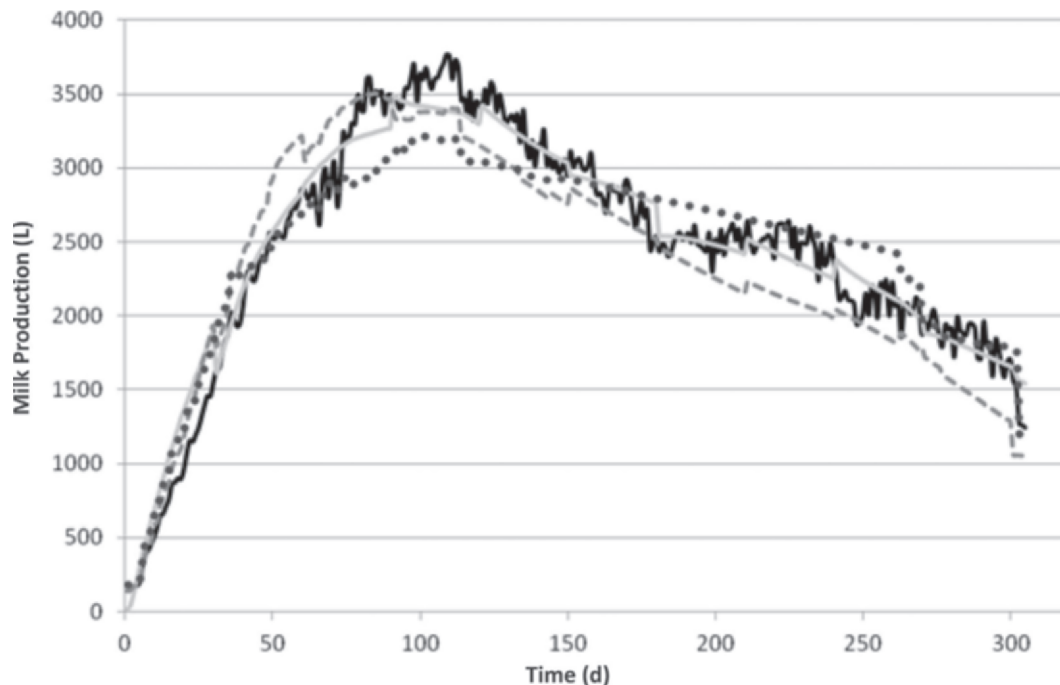


Figure 5. Milk-production forecast for a 30-d moving piecewise horizon. Black solid line = actual milk production, gray solid line = nonlinear auto-regressive model with exogenous input forecast, dashed line = static artificial neural network forecast, dotted line = multiple linear regression forecast.

Table 3. The 10-d moving piecewise horizon forecast error

Range	NARX ¹			SANN ²			MLR ³		
	MAE ⁴	σ^2 ⁵	RMSE ⁶	MAE	σ^2	RMSE	MAE	σ^2	RMSE
0–10	98.1	1,945.2	107.6	44.6	192.2	46.7	73.6	1,932.9	85.7
10–20	119.8	5,584.8	141.2	99.5	1,427.3	106.5	225.7	2,259.5	230.6
20–30	78.3	990.4	84.4	141.8	1,410.4	146.7	272.2	845.8	273.7
30–40	142.5	10,864.1	176.6	282.0	12,172.8	302.9	284.9	1,178.2	287.0
40–50	51.4	1,917.4	67.5	247.5	2,052.0	251.7	60.2	1,590.2	72.2
50–60	45.4	842.4	53.9	331.0	2,136.3	334.3	70.6	2,562.6	86.9
60–70	178.5	12,486.0	210.6	373.2	17,092.0	395.4	100.9	1,975.6	110.2
70–80	172.7	6,721.0	191.1	257.3	31,186.7	312.1	300.8	28,105.3	344.4
80–90	69.3	4,380.0	95.8	68.2	2,829.7	86.5	486.1	9,286.2	495.5
90–100	107.3	9,606.7	145.3	156.6	6,703.6	176.7	370.1	5,725.6	377.7
100–110	108.9	8,266.5	141.8	312.6	12,132.2	331.4	449.5	5,401.0	455.5
110–120	76.5	2,525.6	91.5	283.0	24,882.1	324.0	407.4	5,387.6	413.9
120–130	109.3	5,458.7	132.0	306.9	5,468.0	315.7	422.7	6,719.1	430.6
130–140	81.6	3,148.2	99.1	166.5	11,372.6	197.7	256.0	11,967.3	278.4
140–150	108.7	2,238.6	118.5	184.5	4,399.6	196.0	120.7	7,666.2	149.1
150–160	99.6	2,615.4	111.9	216.6	4,378.2	226.5	122.2	3,086.7	134.2
160–170	61.2	1,668.0	73.6	109.4	4,390.4	127.9	55.6	2,155.4	72.4
170–180	183.8	9,348.1	207.6	123.5	6,722.9	148.2	166.2	11,528.5	197.9
180–190	67.3	2,186.9	82.0	76.5	1,383.1	85.1	276.4	3,708.7	283.0
190–200	65.6	2,509.9	82.6	140.0	4,420.5	155.0	229.3	5,972.6	242.0
200–210	73.7	1,294.9	82.0	282.2	6,362.2	293.3	147.3	5,246.2	164.2
210–220	57.6	1,370.4	68.5	337.7	3,727.2	343.2	124.8	3,860.4	139.4
220–230	199.3	5,913.2	213.6	424.5	4,593.7	429.9	63.6	3,723.8	88.1
230–240	114.7	3,391.7	128.6	400.7	8,242.9	410.9	109.9	6,835.5	137.5
240–250	159.2	2,607.8	167.2	102.1	4,112.3	120.6	403.3	7,158.3	412.0
250–260	135.0	7,185.2	159.4	285.4	12,943.8	307.3	308.3	9,756.1	323.8
260–270	107.6	6,200.3	133.4	233.2	8,764.8	251.3	209.8	9,681.1	231.7
270–280	160.7	5,378.8	176.7	275.3	4,585.4	283.5	58.2	748.2	64.3
280–290	194.3	4,864.1	206.4	323.7	9,467.8	338.0	97.1	1,085.4	102.6
>290	170.1	6,415.9	188.0	254.9	10,837.9	275.3	94.1	4,897.4	117.3

¹NARX = nonlinear auto-regressive model with exogenous input.²SANN = static artificial neural network model.³MLR = multiple linear regression model.⁴MAE = mean absolute error.⁵ σ^2 = variance.⁶RMSE = root mean square error.

horizons due to the updated regression coefficients, but it did not result in any substantial increase in accuracy.

The NARX and SANN were far more successful than the MLR at predicting peak milk production. This attribute is particularly useful for milk processors, because accurate peak milk-supply forecasts are needed to calculate required processing capacity.

The NARX model's ability to dynamically adapt and alter its projected forecast gave it a distinct advantage over other model types for short- to medium-term predictions. Other easily definable herd inputs such as cow parity and culling rate could be added to the NARX with little complication. The model's accurate short-term predictions would make it a useful decision support tool for farm managers over short periods.

NARX Limitations

One limiting factor of the NARX ability to retain information in its short-term embedded memory is

that it is vulnerable to stochastic disturbances (mainly changes in weather) in the time series. These disturbances create noise in the DHMY time series, and in turn this creates outliers. If the end of the preceding horizon contains outliers, this information will be retained in the model's memory and can amplify the divergence between the models trajectory and the actual DHMY series. The specific point of refresh along the times series has a significant influence on the model's prediction accuracy; this phenomena can lead to disparities between predictions in overlapping ranges. This characteristic could possibly retard the model's ability to adapt to real trends. This problem could be reduced with a shorter forecast horizon.

Comparisons with Previous Studies

The farm conditions (stocking rate, calving season) did not differ from the training data, and no dramatic disturbances occurred (disease outbreak, freak weather

Table 4. Totalized prediction error over various forecast horizons

Horizon	NARX ¹		SANN ²		MLR ³	
	RMSE ⁴ (%)	R ²	RMSE (%) ⁵	R ²	RMSE (%) ⁵	R ²
305 d	8.59	0.936	12.03	0.889	10.62	0.917
50 d	8.10	0.941	12.15	0.887	10.68	0.914
30 d	6.77	0.964	11.74	0.894	10.62	0.915
10 d	5.84	0.968	10.70	0.911	10.54	0.916

¹NARX = nonlinear auto-regressive model with exogenous input.

²SANN = static artificial neural network model.

³MLR = multiple linear regression model.

⁴RMSE = root mean square error; RMSE is expressed as a percentage of the average actual milk production.

event) during the target season; this ensured a level of accuracy of $R^2 \geq 0.89$ for each model over the 305-d horizon. Olori et al. (1999) stated that a model's prediction of total herd milk production was good if $R^2 \geq 0.7$, fair if $0.7 > R^2 \geq 0.4$, and poor if $R^2 < 0.4$. Quinn et al. (2005) applied several existing formulas (Wood, 1967; Ali and Schaeffer, 1987; Wilmink, 1987; Guo, 1995) to data originating from an Irish dairy herd. The models performed reasonably well, obtaining R^2 values of 0.63, 0.68, 0.60, and 0.64, respectively. Killen and Keane (1978) applied Wood's lactation curve to Irish milk-production data and obtained R^2 values in the order of 0.7. Olori et al. (1999) achieved an R^2 of 0.94 using the standard lactation curve method to predict

mean weekly herd yield for a group of HF cattle in the United Kingdom. For predicting individual cow yield, the average R^2 was 0.67. The reduction in accuracy was in part due to nontypical lactations. Grzesiak et al. (2006) obtained R^2 values ranging from 0.45 to 0.62 and RMSE of 17.58 to 18.45% with Wood's lactation curve, whereas their ANN model achieved an R^2 of 0.77 and a RMSE of 14.74% for the same data set.

An ANN model for milk-yield prediction using more easily accessible training data was addressed by Grzesiak et al. (2006). In that study the model inputs used were percentage of HF genes, age at calving in months, the month of calving, the DIM after calving, and the lactation number. This model proved to be more ac-

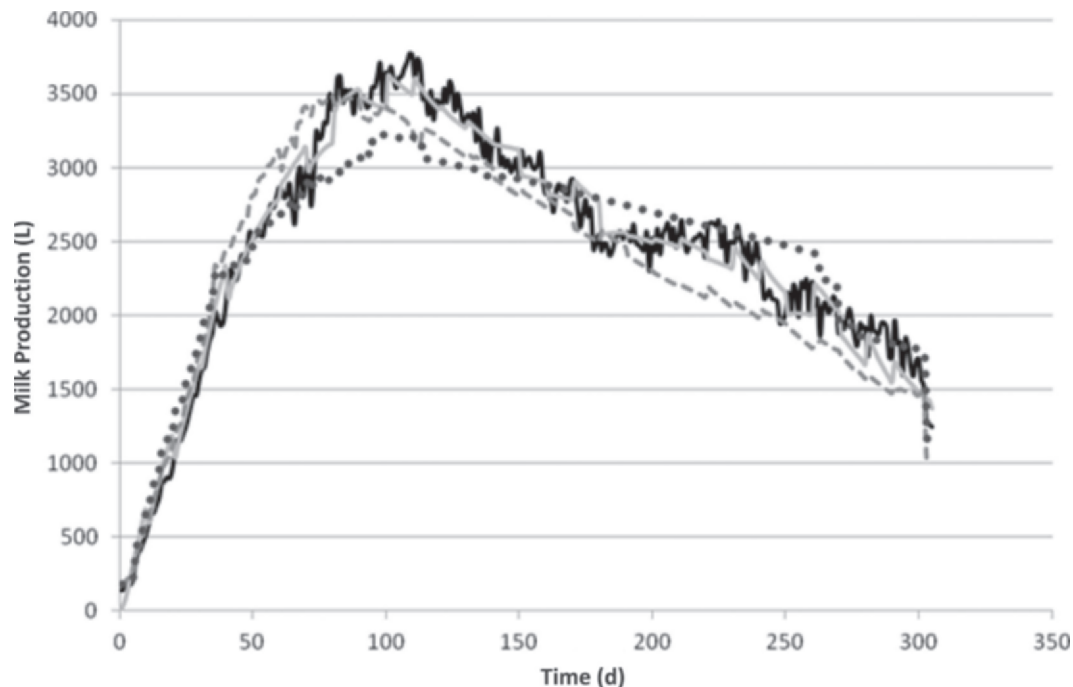


Figure 6. Milk-production forecast for a 10-d moving piecewise horizon. Black solid line = actual milk production, gray solid line = non-linear auto-regressive model with exogenous input forecast, dashed line = static artificial neural network forecast, dotted line = multiple linear regression forecast.

curate than Wood's lactation curve and much easier to implement in comparison with a large regression model; nevertheless, this model still required specific information on each individual cow and would not be possible to apply to a dairy farm without the use of a sophisticated computerized milk recording apparatus.

The SANN and MLR tested in this study produced similar levels of forecast error with the MLR being more accurate for 53.2% of the predicted ranges and having a lower totalized prediction error (Table 4) for each moving horizon. The MLR and SANN used by Grzesiak et al. (2003) achieved R^2 of 0.87 and 0.93 and RMSE of 9.84 and 9.05%, respectively. Sharma et al. (2007) found a prediction accuracy of 92.03% for their SANN and 91.38% for their MLR. Dongre et al. (2012) SANN and MLR models attained prediction accuracies of 86.08 and 85.16%, respectively. The models used in the above studies, however, did use more input variables than the models used in this work.

The predictions of Jones's (1997) empirical Bayes method for fitting Wood's curve produced correlation values with the actual production figures of up to 0.81. Baudracco et al. (2012) applied a mechanistic milk yield model to predict milk production of a pasture-based dairy cow. The model forecast achieved a concordance correlation coefficient of 0.76. The auto-regressive models developed by Macciotta et al. (2002) and Vasconcelos et al. (2004) obtained correlation values with their corresponding data of up to 0.85 and 0.84, respectively, but it is difficult to compare these models to the NARX in this study, because they predicted individual lactation cycles using various test-day figures from incomplete lactations.

It is difficult to compare the effectiveness of different models over individual studies. Each study mentioned above used case-specific data to predict milk supply for a particular horizon for an exact cow or herd in unique conditions. The idiosyncratic nature of milk supply modeling only allows us to make broad comparisons between our results and other studies. Definite conclusions on model accuracy for milk-production forecasting can only be made between the 3 models tested in this body of work.

CONCLUSIONS

The 3 models tested in this study produced daily herd milk-production predictions for 4 different piecewise moving horizons using limited amounts of data (R^2 values from 0.936 to 0.968 for the NARX model, 0.889 to 0.911 for the SANN model, and 0.917 to 0.916 for the MLR model). The NARX was shown to be the most effective milk-production model. It was more accurate than the SANN and MLR for each moving

horizon and over the majority of individual prediction ranges (63.8% of ranges). From this research, the NARX model appears to be a viable alternative to conventional regression models; however, more work must be undertaken to assess the NARX aptitude for milk-production prediction on a large number of commercial dairy farms using different input data. Future work will focus on testing more models at cow and herd level for multiple dairy farms. Factors such as climatic conditions, seasonal calving, and cow parity are also matters of interest in the future.

REFERENCES

- Adeiran, S. A., D. A. Ratkowsky, D. J. Donaghy, and A. E. O. Malau-Aduli. 2012. Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. *J. Dairy Sci.* 95:5344–5356.
- Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Anim. Sci.* 67:637–644.
- Baudracco, J., N. Lopez-Villalobos, C. W. Holmes, E. A. Comeron, K. A. Macdonald, T. N. Barry, and N. C. Friggens. 2012. e-Cow: An animal model that predicts herbage intake, milk yield and live weight change in dairy cows grazing temperate pastures, with and without supplementary feeding. *Animal* 6:980–993.
- Brun-Lafleur, L., L. Delaby, F. Husson, and P. Faverdin. 2010. Predicting energy \times protein interaction on milk yield and milk composition in dairy cows. *J. Dairy Sci.* 93:4128–4143.
- Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. *J. Dairy Sci.* 83:2683–2690.
- Dan Foresee, F., and M. T. Hagan. 1997. Gauss-Newton approximation to Bayesian learning. Pages 1930–1935 in *Proc. IEEE Int. Conf. Neural Netw. Inst. of Electrical and Electronics Eng. (IEEE)*, New York, NY.
- Demuth, H. B., M. T. Hagan, and M. Beale. 2010. *Users' Guide for the Neural Network Toolbox for Matlab*. The Mathworks, Natick, MA.
- Diaconescu, E. 2008. Prediction of chaotic time series with NARX recurrent dynamic neural networks. Pages 248–253 in *Proc. 9th WSEAS Int. Conf. Automation Info.* World Sci. Eng. Acad. Soc., Bucharest, Romania.
- Dongre, V. B., R. S. Gandhi, A. Singh, and A. P. Ruhil. 2012. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livest. Sci.* 147:192–197.
- Grzesiak, W., P. Blaszczyk, and R. Lacroix. 2006. Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric.* 54:69–83.
- Grzesiak, W., R. Lacroix, J. Wójcik, and P. Blaszczyk. 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Can. J. Anim. Sci.* 83:307–310.
- Guo, Z. S. H. 1995. Modelling of the Lactation Curve as a Sub-model in the Evaluation of Test Day Records. *INTERBULL*, Prague, Sveriges Lantbruksuniversitet, Uppsala, Sweden.
- Hagan, M. T., and M. B. Menhaj. 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5:989–993.
- Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ.
- Hecht-Nielsen, R. 1989. Theory of the backpropagation neural network. Pages 593–605 in *Proc. Int. Joint Conf. Neural Netw.* IEEE Press, New York, NY.
- Jones, T. 1997. Empirical Bayes prediction of 305-day milk production. *J. Dairy Sci.* 80:1060–1075.

- Killen, L., and M. Keane. 1978. The shape of lactation curves in Irish dairy herds. *Isr. J. Agric. Res.* 17:267–282.
- Kolver, E. S., and L. D. Muller. 1998. Performance and nutrient intake of high producing Holstein cows consuming pasture or a total mixed ration. *J. Dairy Sci.* 81:1403–1411.
- Lacroix, K. M. W., R. Kok, and J. F. Hayes. 1995. Prediction of cow performance with a connectionist model. *Am. Soc. Agric. Biol. Eng.* 38:1573–1579.
- Lapedes, A., and R. Farber. 1987. *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling*. Natl. Lab., Los Alamos, NM.
- Macciotta, N. P. P., D. Vicario, G. Pulina, and A. Cappio-Borlino. 2002. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *J. Dairy Sci.* 85:3107–3114.
- MacKay, D. J. C. 1992. Bayesian interpolation. *Neural Comput.* 4:415–447.
- McCarthy, B., L. Delaby, K. M. Pierce, F. Journot, and B. Horan. 2011. Meta-analysis of the impact of stocking rate on the productivity of pasture-based milk production systems. *Animal* 5:784–794.
- Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* 58:55–63.
- Ptak, E., and L. R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34:23–34.
- Quinn, N., L. Killen, and F. Buckley. 2005. Empirical algebraic modelling of lactation curves using Irish data. *Ir. J. Agric. Food Res.* 44:1–13.
- Salehi, F., R. Lacroix, and K. M. Wade. 1998. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Comput. Electron. Agric.* 20:199–213.
- Shalloo, L., P. Dillon, M. Rath, and M. Wallace. 2004. Description and validation of the Moorepark dairy system model. *J. Dairy Sci.* 87:1945–1959.
- Sharma, A. K., R. K. Sharma, and H. S. Kasana. 2006. Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. *Neural Comput. Appl.* 15:359–365.
- Sharma, R. K., A. K. Sharma, and H. S. Kasana. 2007. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Appl. Soft Comput.* 7:1112–1120.
- Smith, L. P. 1968. Forecasting annual milk yields. *Agric. Meteorol.* 5:209–214.
- Tekerli, M., Z. Akinci, I. Dogan, and A. Akcan. 2000. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. *J. Dairy Sci.* 83:1381–1386.
- Lin, T., B. G. Horne, P. Tino, and C. L. Giles. 1996. Learning long-term dependencies in NARX recurrent neural networks. *IEEE Trans. Neural Netw.* 7:1329–1338.
- Vasconcelos, J., A. Martins, M. F. Petim-Batista, J. Colaço, R. W. Blake, and J. Carvalheira. 2004. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *J. Dairy Sci.* 87:2591–2598.
- Weigend, A. S., D. Huberman, and D. E. Rumelhart. 1990. Predicting the future: A connectionist approach. *Int. J. Neural Syst.* 13:193–209.
- Wilmink, J. B. M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335–348.
- Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165.